Implementing Distinctive Behavior for Conversational Agents

Maurizio Mancini[1] and Catherine Pelachaud[2]

[1] University of Paris 8
and
DIST Lab, University of Genova, Italy
[2] INRIA - France
and
CNRS, TELECOM - ParisTech, Paris, France

Abstract. We aim to define conversational agents exhibiting distinctive behavior. To this aim we provide a small set of parameters to allow one to define behavior profiles and then leave to the system the task of animating the agents. Our approach is to manipulate the behavior tendency of the agents depending on their communicative intention and emotional state. In this paper we define the concepts of *Baseline* and *Dynamicline*. The Baseline of an agent is defined as a set of fixed parameters that represent the personalized agent behavior, while the Dynamicline is a set of parameters that derive both from the Baseline and the current communicative intention and emotional state.

1 Introduction

We present a model for the definition of conversational agents exhibiting *distinctive behavior*. It means that even if the communicative intentions or emotional states of two agents are exactly the same they behave in a different way according to their global behavior tendencies. Differences are noticeable both in the signals chosen by the agents to communicate and in the quality of their behaviors.

Human communication involves verbal and non-verbal behaviors. People communicate through several modalities like face, gestures, posture. The same beliefs and goals can be communicated in different personalized ways, and the interaction with other people is always influenced by personal behavior tendencies. For example we are able to give a quick but usually precise definition of the global behavior tendency of a person: we say "this person *never* moves, she is expressiveless" or "she is very expansive and gestures a lot while talking" and so on. Some people may tend to use one modality more than the others, other people may use more than one modality at the same time and so on. For example let us consider a person that gestures a lot while speaking. This is a basic trait of that person, we can expect that in most circumstances she prefers to convey non-verbal signals more on the gesture modality than with other modalities. That is the idea we want to capture with the concept of *Baseline* for virtual agents. On the other hand there can be some events or situations in which one's basic tendencies change, and one gestures in a greatly different way. For example

M. Sales Dias et al. (Eds.): GW 2007, LNAI 5085, pp. 163–174, 2009.

a person that never does hand/body gestures while she talks may change her behavior if she is very angry at someone. We embody the current tendency of behavior with the concept of *Dynamicline* of a virtual agent.

In the next Section we briefly describe other systems that have implemented agents exhibiting distinctive behavior. In Section 3 we illustrate the definition of the modality preference and the expressivity parameters. In Sections 4 and 5 we explain how Baseline and Dynamicline are used to generate the final agent's behavior. Then we give an example of our system and we conclude the paper.

2 State of the Art

Several researchers addressed the problem of defining conversational agents exhibiting distinctive behavior. Some of them applied psychological theories to the creation of models that simulate personality, mood and emotion [3]. Others defined parameters that aim to modify the quality of movement dynamically to increase the expressivity of the agent [4,12,14]. In other systems, the agent's behavior, previously stored in script files manually or statistically computed, is selected during the interaction with the user [2,10].

Ruttkay et al. [15,18] propose the idea of behavior style, defined in terms of *when* and *how* the agent uses certain gestures. Styles are implemented by selecting gestures from a *style dictionary* that defines both which gestures an agent has in its repertoire and its habits in using them. The style dictionaries are written in GESTYLE. This language specifies which modalities should be used to display non-verbal behaviors and is also used to annotate the text that the agent has to utter. Ball and Breese [3] propose a model for individualization for virtual agents in which the final behavior is computed depending on the agent's current emotional state and personality by choosing the most appropriate style. The PAR model of Allbeck et al. [1] offers a parameterization of actions. The actions that the agent is able to carry out are defined together with the conditions that need to be true in order to perform the actions. Conditions can refer to the state of other agents or objects in the agent's environment. EMOTE [4] is a system for creating differentiated gestures. Starting from Laban's annotation scheme, gestures are defined in terms of *Effort* and *Shape*. Effort gives information about the sense of impact (delicate vs strong), speed (indulging vs urgent/sudden) and control (uncontrolled/abandoned vs controlled/tense) of movement. space, weight, time and flow. Shape defines the movement path in the 3D space. Neff et al. [14] found out some key movement properties by reviewing arts and literature, for example from theater and dance. They found that body and movement characteristics such as balance, body silhouette (contour of the body), position of torso and shoulder, etc. influence the way in which people perceive the others. They implemented three motion properties into animated characters: the pose of the character, the timing of movements and the transition from one pose to another. In this model physical constraints are also considered, for example gravity and body balance, to obtain a very realistic animation. In André et al. [2] the agent's behavior depends both on a script that describes what the agent has to communicate to the user (for example how to do a reservation for a room in a hotel's website) and its personalized behavior. The last one includes idle movements like for example tapping with its foot while the user does nothing or jumping when the mouse passes over the agent's figure. Maya et al. [12] define a taxonomy to classify the influences on people's behavior. Intrinsic influences are personal habits that derive from personality, culture, gender, etc. Contextual influences come from the physical environment in which the person is acting. Finally dynamic influences represent the person's emotional states, beliefs and goals during the conversation. In our model we also use the notion of dynamic influence but we look at how these influences can be computed in term of movement quality and modalities preference without looking at the factors that could have caused them. In Michael Kipp's work [9] the author presents a gesture animation system based on statistical models of human speaker's gestures. Videos of interviewed people have been manually annotated in terms of gestures types (iconic, deictic, etc. [13]), together with their frequency of occurrence and timing (that is the synchronization between the gesture stroke and the most emphasized syllable of the sentence). The statistics on the speaker's gestures are then used to model the agent's set of preferred gestures (the probability of their occurrence is computed from the annotated gesture frequency) and synchronization tendency (for example an agent can perform gesture strokes always synchronized with speech emphasis). In a more recent work [10], the agent's gestures selection can be human authored or automatically learned using machine learning algorithms on the basis of previously annotated scripts. In our work we look at behavior qualitative differences rather than gesture types differences. Kipp's approach and ours are thus complementary. Similarly to our work, Kipp does not model the possible causes of visible variations in behavior.

3 Modality Preference and Behavior Expressivity

Conversational agents are graphical representations of humans that are increasingly used in a large variety of applications to help, assist or direct the user in performing a wide range of tasks. They can communicate to the user multimodally, that is by using many modalities at the same time. In our work, agents produce signals on the following modalities:

- face (eyebrows/eyelids/mouth/cheek movements)
- head movement (head direction and rotation, such as nods and shakes)
- gestures (arms and hands movements)
- body posture (upper part of the body movements)

People differ in the way they use their modalities: one can be very expressive on the face, another can gesture a lot. The concept of *Modality preference* encompasses this variability in the modalities use. People can also differ in the quality of their behavior. For example, one can have the tendency to do large hand gestures at a fast pace. Thus behavior expressivity is also a characteristic of an agent.

3.1 Modality Preference

People can communicate by being more or less expressive in the different modalities. The *modality preference* represents the agent's degree of preference for each available modality. If for example we want to specify that the agent has the tendency to mainly use hand gestures during communication we assign a high degree of preference to the *gesture* modality, if it uses mainly the face, the face modality is set to a higher value, and so on. For every available modality (face, head movement, gesture, posture), we define a value between 0 and 1 which represents its degree of preferability. Agents can also use two or more modalities with the same degree of preference. This means that the agent communicates with these modalities equally.

3.2 Behavior Expressivity

Behavior expressivity is an integral part of the communication process as it can provide information on the current emotional state, mood, and personality of the agent [20]. We consider it as the manner the physical behavior is executed. Starting from the results reported in [20], we defined and implemented [7,8] a set of parameters that affect the qualities of the agent's behavior such as its speed, spatial volume, energy, fluidity. Thus, the same gestures or facial expressions are performed by the agent in a qualitatively different way depending on the following parameters:

- *Overall activation (OAC)*: amount of activity (quantity of movement) across several modalities during a conversational turn (e.g., simultaneous use of facial expression, gaze, gesture to visualize communicative intentions passive/static or animated/engaged).
- *Spatial extent (SPC)*: amplitude of movements (e.g., extension of the arms; amplitude of eyebrow raise)
- *Temporal (TMP)*: duration of movements (e.g., quick versus sustained actions)
- *Fluidity (FLD)*: smoothness and continuity of overall movement (e.g., smooth, graceful versus sudden, jerky)
- *Power (PWR)*: dynamic properties of the movement (e.g., weak/relaxed versus strong/tense)
- *Repetitivity (REP)*: tendency to rhythmic repeats of specific movements.

4 Baseline and Dynamicline

In our model for conversational agents exhibiting distinctive behavior we want to capture the idea that people have tendencies that characterise globally their behavior, but these tendencies can change in situations arising after some particular events. We introduce the concepts of *Baseline* and *Dynamicline*, which both contain information on the agent's modalities preferences and expressivity but with different time span: while the Baseline is the *global* definition of how the agent behaves in most situations, the Dynamicline is the *local* specification of the agent's behavior (for example during a given agent's emotional state).

In our model, Baseline and Dynamicline do not only differ by their meaning (global vs local behavior tendency) but also by the fact that the Baseline is an input parameter, that is, it is used to define some characteristics of an agent, while the Dynamicline is automatically computed by the system at runtime, depending on the agent's current communicative intention or emotional state.

4.1 Baseline

We define the Baseline by the couple *(Mod,Expr)* where:

- *Mod*: is the modality preference. As described in Section 3.1 this is the agent's tendency to use its modalities during communication. The modality preference is defined by assigning a real value between 0 and 1 to each modality.
- *Expr*: is the behavior expressivity. It is a set of expressivity values (see Section 3.2) that represents the base behavior tendency of the agent. An agent could for example tend to do slow and smooth gestures while another agent could tend to move in a fast and jerky manner. Note that we implemented expressivity separately for each modality, that is there is a set of expressivity parameters for each modality.

4.2 Behavior Quality Computation

Figure 1 outlines the module called *Behavior Quality Computation*, or *BQC*, which computes the agent's Dynamicline at runtime. The data provided as input is the agent's Baseline, which is a constant, and its current communicative intention or emotional state. Each time the communicative intention or emotional state varies the module computes a new Dynamicline for the agent.

During the execution of the BQC module, the modality preference and the expressivity parameters contained in the Baseline are modulated depending on the agent's actual communicative intention and the resulting values are stored in the Dynamicline. It means that communicative intention has a different impact on the Dynamiclines of two agents having different Baselines. For example, if an agent has a general tendency (Baseline) to perform movements with average speed/amplitude and to use hand gestures moderately then in a sad state it does very few hand gestures with very low amplitude and speed. On the other hand, an agent with a general tendency of gesturing a lot with fast and large movements even when being sad continues making gestures that would be fewer in number and with a lower expressivity (average speed and amplitude).

As shown in Figure 1, the computation of the agent's Dynamicline consists in "applying" some *behavior qualifiers* to the agent's Baseline. The behavior qualifiers allow us to define mathematical operations to be performed over the parameters contained in the agent's Baseline each time a certain intention must be communicated.

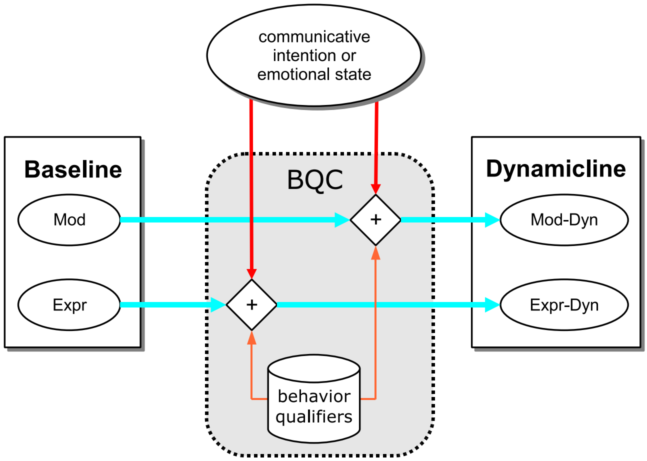

Fig. 1. The agent's Baseline, communicative intention or emotional state determine the computation of the agent's Dynamicline

For example, we could define a behavior qualifier that represents the following description:

"a angry state (i) increases the degree of bodily activation and at the same time (ii) the speed, amplitude and energy of movements are very high".

To implement such qualifiers we defined an XML-based language in which we could represent:

- the name of the parameter of the Dynamicline affected by the qualifier;
- the mathematical operator that is used in the qualifier (addition, subtraction, maximum and so on);
- the name of the parameter of the Baseline involved in the computation specified by the qualifier;

So the qualifier in the above example could be described with the following pseudo-code:

```
behavior qualifier for anger:
      affects body OAC by multiplying by a factor;
      affects TMP, SPC and PWR of all modalities by setting them to
      high values;
```

In the BQC module (see Figure 1), a repository of behavior qualifiers is stored, that is, a behavior qualifier for each possible agent's communicative intention or emotional state is defined. For every variation of one of these two data in input the module computes the agent's Dynamicline by selecting the corresponding behavior qualifier from the repository. The modulations induced by the behavior sets on the agent's Baseline to compute the Dynamicline have been defined starting from the studies of Wallbott [19] and Gallaher [6] and have been tested in previous works [7,11].

4.3 Dynamicline

The output of the BQC module (see Figure 1) is the agent's current Dynamicline. Structurally the Dynamicline is identical to the Baseline and is modeled by the couple *(Mod-Dyn,Expr-Dyn)* where:

- *Mod-Dyn*: The agent's current modality preference. It represents the agent's tendency to use its modalities given a certain communicative intention or emotional state. It is obtained by modulating the modality preference *Mod* contained in the Baseline depending on the current communicative intention or emotional state.
- *Expr-Dyn*: The agent's current expressivity parameters. It represents the agent's expressivity of movements given a certain communicative intention or emotional state. It is obtained by modulating the expressivity parameters *Expr* contained in the Baseline depending on the current communicative intention or emotional state.

5 Distinctive Behavior Generation

The presented work has been implemented within the Greta conversational agent framework [16]. The input of such a system is a text file in which both text to be spoken and the associated communicative intentions and emotional states are written. The text is translated into a speech file by a speech synthesizer while another process computes the animation of a virtual human that performs facial, arm and body gestures. Figure 2 represents the architecture of our system, showing how the modules are connected.

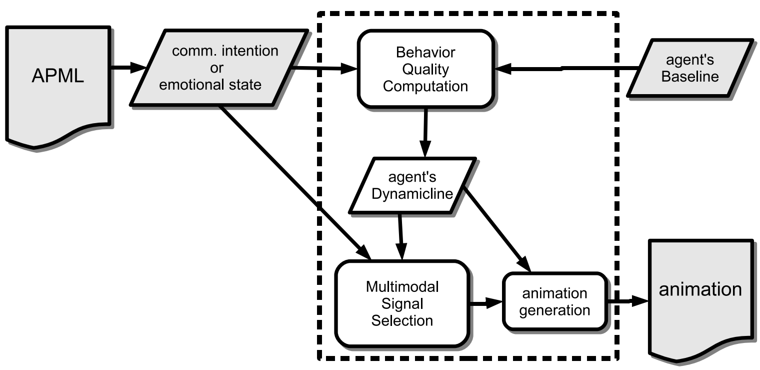

Fig. 2. The architecture of our system for distinctive behavior

```
01 <apml>
02
03 <rheme affect="anger">
04 <emphasis deictic="point">
05 what are you doing
06 </emphasis>
07 </rheme>
08
09 <performative type="warn">
10 Did you
11 <emphasis>
12 pay
13 </emphasis>
14 attention to what I said or
15 <emphasis deictic="self">
16 not
17 </emphasis>
18 </performative>
19
20 </apml>
```

Fig. 3. An APML file

Here is a description of each system component:

- *APML*: The input of our system is a file with a high-level description of the communicative intentions that the agent aims to communicate. It is written in APML [5], an XML-based language whose tags represent the communicative intentions [17]. For example in Figure 3 the APML tags surrounding the text specify that the agent is going to *warn* the user, see the *performative* tag from line 09 to line 18. The emotional state of the agent is also specified with the *affect* attribute of the *rheme* tag, from line 03 to line 07.

  The input APML file is read sequentially, and the contained tags are sent one by one to the behavior generation module. In Figure 2 the block called *communicative intention or emotional state* contains one APML tag at a time.
- *Behavior Quality Computation*: Depending on the current APML tag which can be the communicative intention or emotional state of the agent, this module computes the agent's Dynamicline from the Baseline, as explained in Section 4.
- *Multimodal Signal Selection*: The Dynamicline of the agent and the current APML tag are then used to decide which signals (facial expressions, gestures, etc) have to be performed by the agent. Starting from the current APML tag, the system looks in a *multimodal lexicon* that associates communicative intentions with possible combinations of multimodal signals. Moreover, the multimodal lexicon defines some constraints for the signals selection, as for example using always one (or more) specific modality to convey the given meaning. Then the modality preference of the agent (see Section 3.1) computed in the Dynamicline is used to choose the set of multimodal signals with the highest preference. For example if the current communicative intention contained in the APML file can be communicated with two or more groups of signals, the system chooses to generate the signals that are from

the modalities with the highest preference. If these modalities cannot be instantiated (because, for example, the modalities are already in use) other signals with lower preference is used.

– *animation generation*: This is the last step of the generation process. Once all the signals corresponding to a given APML tag have been instantiated, this module computes the animation of the agent. The set of expressivity parameters contained in the agent's Dynamicline are now applied to the agent's movements [8].

6 A Working Example

We aim to define conversational agents that, even if they have the same communicative intention and emotional state, they would exhibit different multimodal behavior. We saw how the Baseline can be instantiated into the Dynamicline and how the contained data could influence the selection of the agent's communicative modalities and the expressivity of behaviors.

Let us see an example of how the proposed system works in practice. We want to model two conversational agents, *agent1* and *agent2*, with different behavior Baselines (reported in Figure 4). To simplify the explanation, we consider only two communication modalities: facial expressions and arm/hand gestures.

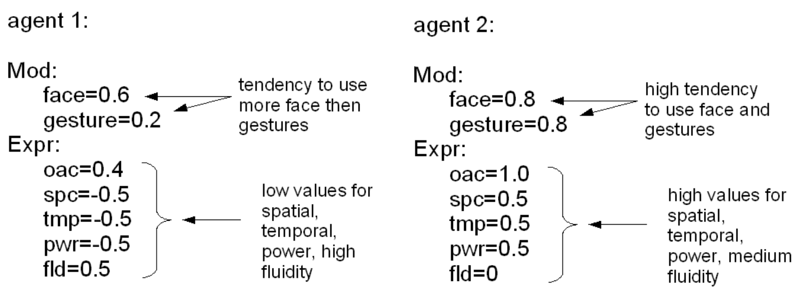

Fig. 4. Examples of Baseline: an *indolent*-type agent on the left; an *exuberant*-type agent on the right

Agent 1 is defined by a Baseline with a very low preference for doing gestures. It prefers communicating through facial expressions. It is not very expressive and its overall behavior is slow, powerless and low in amplitude. On the other hand, agent 2 uses equally its face and gestures. It is very expressive and uses a lot of non-verbal behaviors during communication. Its overall behavior is expansive, rapid and powerful.

Let us see how agent 1 and agent 2 differ in communicating the utterances of the APML example reported in Figure 3.

For the tag *affect="anger"* the system looks in the multimodal lexicon and finds that the tag corresponds to a facial expression (frown + tense lips). The tag spans from lines 3 to 7 and contains an *emphasis* tag (lines 4-6) that can be

	agent 1	*agent 2*
start	Baseline: • Emo: neutral • Mod: face=0.6, gesture=0.2 • Expr: oac=0.4, spc=-0.5, tmp=-0.5, pwr=-0.5, fld=0.5	Baseline: • Emo: neutral • Mod: face=0.8, gesture=0.8 • Expr: oac=1.0, spc=0.5, tmp=0.5, pwr=0.5, fld=0
	tag: affect+emphasis	***tag: affect+emphasis***
step 1	Dynamicline: • Mod: face=0.7, gesture=0.4 • Expr: oac=0.6, spc=-0.1, tmp=-0.1, pwr=-0.1, fld=0.1 Selected signals: • face=anger, gesture=nothing	Dynamicline: • Mod: face=0.9, gesture=1.0 • Expr: oac=1.0, spc=1.0, tmp=1.0, pwr=0.9, fld=-0.4 Seleced signals: • face=anger, gesture=beat
	tag: performative	***tag: performative***
step 2	Dynamicline: • Mod: face=0.7, gesture=0.6 • Expr: oac=0.6, spc=-0.1, tmp=-0.1, pwr=-0.1, fld=0.1 Selecetd signals: • gesture=deictic(single)	Dynamicline: • Mod: face=0.9, gesture=1.0 • Expr: oac=1.0, spc=1.0, tmp=1.0, pwr=0.9, fld=-0.4 Selected signals: • gesture=deictic(with repetition)

Fig. 5. Evolution of the Baselines of an *indolent* and an *exuberant* agent. There are differences in the values of the expressivity parameters and the chosen signals.

performed by a combination of signals on the face (raised eyebrows) and gesture (beat). At this point the Dynamiclines of the two agents are computed by the system. The current communicative intention (emphasis) and emotional state (anger) are used to modulate the agents' Baselines (see Figure 5, line *start*) and to obtain the corresponding Dynamiclines (see Figure 5, line *step 1*). Based on the computed Dynamiclines, agent 1 assumes an angry facial expression, doing no gesture (the gesture modality in the agent's modality preference is still low). Moreover, the angry face is not very intense (spatial parameter). Instead agent 2 performs both an angry facial expression and a beat gesture, because the two modalities are high in its Dynamicline. Both the signals are intensified (spatially expanded, fast, powerful) by the agent's expressivity parameters.

The same computation is repeated for each communicative intention in the APML file. As shown in Figure 5 (see line *step 2*), for the *performative* tag (lines 9-18 of Figure 3) the Dynamicline of agent 1 has a quite high degree of preference for the gesture modality, so the agent performs a deictic gesture, even if globally (due to its Baseline) it prefers to do no gesture.

7 Conclusion and Future Work

We have presented a system for creating agents exhibiting *distinctive behavior*. We define the concept of Baseline as the agent's global expressivity and tendency in using communicative modalities. Depending on the agent's Baseline, communicative intention and emotional state, our system computes the agent's Dynamicline, which is then used in the process of behavior selection and generation. The consequence is that we are able to distinguish the behavior of two agents that have been defined with different Baselines, even if their communicative intentions and emotional states are the same.

One of the main limitations of our system concerns the definition of the behavior qualifiers needed to compute the Dynamicline from the Baseline. Defining these qualifiers is neither obvious nor simple, because only few experimental data is available about the variation of human nonverbal behavior depending on the communicative and emotional context. Further research direction we foresee is to automatically extract these behavior qualifiers from videos.

We also aim to add the layering of the effects of communicative intentions and emotional states that happen at the same time. How could we consider the effects of, for example, a communicative intention of "helping the user" happening at the same time of an emotional state of anger? In the future we could allow one to configure these superpositions of communicative intentions and emotional states.

Finally we need to further evaluate the agent's behavior resulting from the computation performed by our system. We need to verify that we can characterize an agent with its behavior tendencies, by investigating if users could recognize a certain agent just by watching its behavior. Another evaluation of the system could consist in comparing the quality of the interaction between users and agents exhibiting or not distinctive behavior.

References

1. Allbeck, J., Badler, N.: Toward representing agent behaviors modified by personality and emotion. In: Workshop on Embodied Conversational Agents - Let's specify and evaluate them! ACM Press, New York (2002)
2. André, E., Rist, T., Müller, J.: Integrating reactive and scripted behaviors in a life-like presentation agent. In: Second International Conference on Autonomous Agents, pp. 261–268 (1998)
3. Ball, G., Breese, J.: Emotion and personality in a conversational agent. In: Prevost, S., Cassell, J., Sullivan, J., Churchill, E. (eds.) Embodied Conversational Characters. MIT Press, Cambridge (2000)
4. Chi, D., Costa, M., Zhao, L., Badler, N.: The EMOTE model for effort and shape. In: Proceedings of the 27th annual conference on Computer graphics and interactive techniques, pp. 173–182. ACM Press/Addison-Wesley (2000)
5. DeCarolis, B., Pelachaud, C., Poggi, I., Steedman, M.: APML, a mark-up language for believable behavior generation. In: Prendinger, H., Ishizuka, M. (eds.) Life-Like Characters. Cognitive Technologies, pp. 65–86. Springer, Heidelberg (2004)
6. Gallaher, P.E.: Individual differences in nonverbal behavior: Dimensions of style. Journal of Personality and Social Psychology 63(1), 133–145 (1992)
7. Hartmann, B., Mancini, M., Buisine, S., Pelachaud, C.: Design and evaluation of expressive gesture synthesis for embodied conversational agents. In: Third International Joint Conference on Autonomous Agents & Multi-Agent Systems, Utretch (July 2005)

8. Hartmann, B., Mancini, M., Pelachaud, C.: Towards affective agent action: Modelling expressive ECA gestures. In: International conference on Intelligent User Interfaces - Workshop on Affective Interaction, San Diego, CA (2005)
9. Kipp, M.: Gesture Generation by Imitation (2005), http://Dissertation.com
10. Kipp, M.: Creativity meets automation: Combining nonverbal action authoring with rules and machine learning. In: Gratch, J., Young, M., Aylett, R.S., Ballin, D., Olivier, P. (eds.) IVA 2006. LNCS, vol. 4133, pp. 230–242. Springer, Heidelberg (2006)
11. Mancini, M., Bresin, R., Pelachaud, C.: From acoustic cues to an expressive agent. In: Gibet, S., Courty, N., Kamp, J.-F. (eds.) GW 2005. LNCS, vol. 3881, pp. 280–291. Springer, Heidelberg (2006)
12. Maya, V., Lamolle, M., Pelachaud, C.: Influences on embodied conversational agent's expressivity: Towards an individualization of the ecas. In: Proceedings of the Artificial Intelligence and the Simulation of Behaviour, Leeds, UK (2004)
13. McNeill, D.: Hand and Mind: What Gestures Reveal about Thought. University Of Chicago Press (1992)
14. Neff, M., Fiume, E.: AER: Aesthetic Exploration and Refinement for Expressive Character Animation. In: Proceedings of the 2005 ACM SIGGRAPH/Eurographics symposium on Computer animation, pp. 161–170. ACM Press, New York (2005)
15. Noot, H., Ruttkay, Z.: Gesture in style. In: Camurri, A., Volpe, G. (eds.) GW 2003. LNCS (LNAI), vol. 2915, pp. 324–337. Springer, Heidelberg (2004)
16. Pelachaud, C.: Multimodal expressive embodied conversational agents. In: MULTIMEDIA 2005: Proceedings of the 13th annual ACM international conference on Multimedia, pp. 683–689. ACM Press, New York (2005)
17. Poggi, I.: Mind markers. In: Trigo, N., Rector, M., Poggi, I. (eds.) Gestures. Meaning and use, University Fernando Pessoa Press, Oporto (2003)
18. Ruttkay, Z., Pelachaud, C., Poggi, I., Noot, H.: Exercises of style for virtual humans. In: Canamero, L., Aylett, R. (eds.) Animating Expressive Characters for Social Interactions, J. Benjamins Publishing(to appear)
19. Wallbott, H.G.: Bodily expression of emotion. European Journal of Social Psychology 28, 879–896 (1998)
20. Wallbott, H.G., Scherer, K.R.: Cues and channels in emotion recognition. Journal of Personality and Social Psychology 51(4), 690–699 (1986)